图书情报工作
第64卷 第16期 2020年8月

ChinaXiv合作期刊

学科前沿分析的方法与实践*

——以北京大学图书馆为例

■ 张春红[1] 肖珑[1] 贺飞[2] 唐勇[1]

[1] 北京大学图书馆 北京 100871 [2]北京大学学科建设办公室 北京 100871

摘 要：[目的/意义] 通过对比已有的前沿分析方法与实践，梳理和总结前沿分析常用方法的利弊及效果，厘清学科前沿的概念，对图书馆开展基于多源数据的学科前沿分析服务进行探索。[方法/过程] 通过文献和案例调研，运用对比分析、案例分析等方法，对已有前沿分析报告的方法与成果进行比较，并重点以北京大学图书馆的《北京大学科学研究前沿》及《中国高校研究前沿》为案例，介绍高校图书馆基于多源数据进行学科前沿分析服务的内涵、特色、指标、研究方法等，为高校图书馆开展学科前沿分析服务提供借鉴。[结果/结论] 高校图书馆的学科前沿分析服务应精准对标学科发展需求、多方位获取和多维度挖掘数据，综合运用前沿分析的各种方法，明确各学科具有竞争优势的研究领域和研究方向，评估和判断各学科研究的未来发展态势，为一流学科建设的布局及资源投入提供参考。

关键词：前沿分析 前沿报告 学科研究前沿 热点分析 共词分析
分类号：G250
DOI：10.13266/j.issn.0252-3116.2020.16.005

chinaXiv:202304.00133v1

探索分析中国高校各学科的研究前沿具有重要的战略意义。党的十九大报告[1]提出，"创新是引领发展的第一动力，是建设现代化经济体系的战略支撑。要瞄准世界科技前沿，强化基础研究，实现前瞻性基础研究、引领性原创成果重大突破。"国务院印发的《统筹推进世界一流大学和一流学科建设总体方案》[2]提出，高校要"提高基础研究水平，争做国际学术前沿并行者乃至领跑者。推动加强战略性、全局性、前瞻性问题研究，着力提升解决重大问题能力和原始创新能力。"可以看出，梳理各学科研究前沿的现状和分布、明确大学具有竞争优势的基础研究方向、评估和判断大学基础研究的未来发展态势并形成年度的动态分析，正是大学图书馆开展学科前沿分析服务、为大学学科建设与发展战略提供支撑的内涵和价值所在。

1 研究背景和现状分析

1.1 学科前沿的概念亟待厘清

学科前沿相关的概念最早是由 D. J. Price 在 1965 年首次引入的"研究前沿"[3]，近年来引起广泛关注。通过文献调查可以发现，不同的学者和前沿研制的主体对学科前沿这个概念的理解有所不同。从论文引用关系的角度来界定研究前沿的学者包括 H. Small[4]、O. Persson[5]、E. Garfield[6]、S. A. Morris 等[7]；把突发词、研究热点和论文新出现的主题界定为研究前沿的学者包括 J. Kleinberg[8]、C. Chen[9]、黄晓斌和吴高[10]、陈仕吉[11]、张惠荣等[12]。而在实践中，科睿唯安公司与中国科学院（以下简称"中科院"）文献情报中心自 2014 年起连续每年发布的《研究前沿》，从引用关系的角度把研究前沿界定为由一组高被引的核心论文和一组共同引用核心论文的施引文献组成。欧盟网络和信息安全局发布的《隐私和数据保护的设计——从政策到工程》前沿报告把前沿界定为研究热点。可见，前沿尤其是学科前沿的概念亟待厘清。

1.2 多源数据融合是学科前沿研制的趋势

目前学科前沿的识别有以下 3 条路径：

（1）学科专家识别的路径。即特定领域的专家学

* 本系文 2018 年教育部科学事业费重大项目"中国高校研究前沿分析——文献计量学视角"（项目编号：2017ZL17）研究成果之一。

作者简介：张春红（ORCID：0000-0001-5111-4508），研究馆员，硕士；肖珑（ORCID：0000-0003-1923-0626），研究馆员；贺飞（ORCID：0000-0002-9062-4567），研究员，硕士；唐勇（ORCID：0000-0003-2335-4103），副研究馆员，硕士，通讯作者，E-mail：tangyong@lib.pku.edu.cn。

收稿日期：2019-12-25 修回日期：2020-04-05 本文起止页码：36-44 本文责任编辑：王传清

者,通过全面梳理特定学科的研究文献资料,并结合自身的学术积累,分析本学科领域的发展状况和未来发展趋势;或者特定机构利用专家调查问卷或咨询会议等方式,结合德尔菲法等方法,整理提炼不同专家观点,形成代表该领域发展趋势的前沿预测报告,识别当前制约该领域发展的重大基础问题。尽管专家识别的路径高度依赖专家的水平和主观性,在速度、效率和准确性上有一定的局限,但在现有的技术条件下,由于可以充分利用专家的智慧和学科内积累的经验,专家判断法依然是识别和预测科技发展态势的重要手段。如中科院每年发布的《研究前沿》,采用大数据统计和专家鉴定结合的方式;2020 年 1 月发布的第 17 届人文社会科学领域的《2019 年度中国十大学术热点》也是将大数据分析和专家评议相结合进行评选的。

(2)文献计量学的路径。通过跟踪某领域或主题出现新进展时呈现的文献计量特征,以文献为核心对其内部知识结构进行挖掘,对其发展规律和特点进行探测并予以揭示,进而识别该领域或主题的发展趋势和动向。1965 年,D. J. Price[3]把近期发表且被引频次较高的文献定义为科学研究前沿,这是计量学领域对研究前沿探讨的开端。此后,H. Small[4]、O. Persson[5]、E. Garfield[6]、S. A. Morris 等[7]进一步拓展了此概念,这些学者的研究都是基于论文引用关系来识别学科前沿。1988 年,J. Law 等[13]探讨了共词分析法应用于学科热点主题的探测。此后,学者也积极地对基于文本内容的学科前沿识别方法进行了探讨,如,J. Kleinberg[8]提出了考虑词频密度的突发词检测算法。C. Chen[9]提出将一群突现的动态概念及潜在的研究主题定义为研究前沿。A. Aris 等[14]把新出现的学科主题定义为研究前沿。

通过文献调研可以看出,基于文献计量学的路径包括以下几种方法:①基于引用关系的分析方法,此方法按引用类型不同可分为直接引文分析、共引分析和耦合分析 3 种;②基于文本内容的探测方法,包括词频分析法、共词分析法和主题模型法 3 种;③基于引文和文本内容综合分析的探测方法。方法①最大的缺点是滞后性,即一篇文献从出版到被引需要一定的时间。因此很多学者试图从文本内容的角度出发,探索更直接、更有说服力的学科前沿识别方法,即方法②。方法②可以有效克服引文方法的时滞性,且更为直接,但由于词语在不同的语境下所表达的含义可能不尽相同,而且单个词表达的意思不够明确具体,可以说方法②最大的缺点是语义的缺乏。由于单独使用①和②来识别学科前沿都存在一定的局限性,所以许多学者采用方法③,将两种方法结合起来综合使用,希望发挥各个方法的优点,弥补各自的不足。R. Braam 等[15]最早提出将词和共被引结合起来揭示科学研究结构,此后一些学者如 P. V. D. Besselaar 和 G. Heimeriks[16]也进行了相关探讨,但是由于文献计量学是基于成果数据,方法③也未取得完美的结果,所以前瞻性也一直受到质疑。尽管基于文献计量学的方法有一定的局限,但文献计量学方法目前仍然是主流的分析方法。如 2016 年,郑彦宁等[17]探讨了基于共词分析(关键词共现)的研究前沿识别方法;2017 年,M. Li 等[18]对共词分析方法进行改进,用 ARM-KM 模型识别研究前沿;2018 年,T. Zhang 等[19]使用共词分析和引文共引分析方法识别医学大数据领域的研究前沿;J. Hou 等[20]使用共引分析法识别信息科学领域的研究前沿;梁丽等[21]探讨了关键词共现聚类方法、引文共引方法、高频词分析方法、专家分析方法等,认为将各种方法结合起来,相互印证对照,才能够更好地保障研究结果的可靠性和完备性;2019 年,V. Aryadoust 等[22]将共引方法用于识别眼动领域学科前沿;中科院发布的《2019 研究前沿》把一组高被引的核心论文和一组共同引用核心论文的施引论文定义为学科前沿;陈雪等[23]探讨了基于共词分析的国内图书情报领域本体研究热点。

(3)大数据环境下基于多源数据的分析路径。大数据环境下,为了更全面、更客观地识别学科前沿,学者开始探讨利用多源数据进行学科前沿的识别。①基于两种数据源的探讨,如:I. Park 等[24-25]、许晓阳等[26]使用专利和论文数据构建识别能源领域学科前沿的方法;白如江等[27]通过对科技规划文本和基金项目数据文本两种不同数据源中蕴含的科学研究前沿主题相似度对比分析,揭示前沿领域竞争态势;刘博文等[28]从基金项目数据和论文数据融合的角度探讨科学研究前沿主题的识别。②基于 3 种数据源的探讨,如孙震[29]利用科学论文的引用数据、下载使用数据和替代计量数据,构建了研究前沿的集成识别模型。③基于更多数据源的探讨,如:张维冲等[30]使用专利、期刊论文、学位论文、会议论文、图书、基金项目、行业报告共 7 种不同的科技文献类型,来揭示区块链技术的发展情况;谭晓等[31]基于论文、专利、科技规划、资助项目 4 种数据源设计前沿识别模型。可以看出,已有的关于多源数据融合方法的探讨在数据源上仍然是以成果数据为数据源,鲜少融合用户使用数据。而关于融合方法多是基于主题的融合方法。基于多源数据融

chinaXiv:202304.00123v1

ChinaXiv合作期刊

合的学科前沿识别是一个比较新的领域，在数据源的选择、数据融合广度和融合深度上都需要进一步探讨。

鉴于学科专家识别和文献计量学分析等途径和方法都有一定的局限，为了更全面、更科学、更有前瞻性地识别学科前沿，多源数据融合分析是未来学科前沿研制的趋势之一。

1.3 图书馆将会成为学科前沿研制的重要主体

从实践角度来看，也可以发现不同的关于“前沿”的定义，很多机构从不同的角度、运用不同的数据和分析研究方法发布了不同领域的研究前沿报告，比较著名的如科睿唯安公司与中国科学院文献情报中心自 2014 年起连续每年发布的《研究前沿》[32]、中国工程院战略咨询中心与科睿唯安公司共同发布的《全球工程前沿 2018》[33]、麦肯锡公司发布的《人工智能(AI)前沿》报告[34]、联合国环境署发布的《2017 前沿》[35]和《2018/19 前沿：新兴环境问题》[36]报告等(见表 1)。可见，现有前沿分析报告均由专业机构或咨询公司发布；现有前沿分析未能针对特定目标用户，并非需求驱动；不同组织机构发布的前沿分析报告其适用的范围、关注的角度、分析的方法等均有很大的差异。

近年来国家创新战略需求增加，情报分析服务升级，智库服务开始发展，图书馆开始在科研和学科发展支持、学科评估评价等服务框架下，提供各个学科/研究领域的研究前沿分析，比较具有代表性的包括北京大学图书馆的《北京大学科学研究前沿(2018 版)》[37]、上海交通大学图书馆的课题前沿追踪服务、东南大学图书馆的课题咨询与前沿跟踪服务等。

chinaXiv:202304.00133v1

表 1 主要前沿分析方法和形成的报告一览

实践案例	对前沿的界定	分析方法	方法概述	特色
科睿唯安的 ESI 数据库 科睿唯安公司参与发布的《研究前沿》《全球工程前沿》等	研究前沿由一组核心论文组成，这些论文均为高被引论文，且彼此之间具有共被引关系	引文分析法	通过持续跟踪全球最重要的学术论文，研究学术论文被引用的模式及其聚类，特别是成簇高被引论文频繁地共同被引用的情况以发现学科前沿[32,37]	揭示了不同研究者在探究相关科学问题时所产生的某种关联，尽管这些研究人员的背景不同或来自不同的学科领域，引文分析法不是基于单个文献/作者/知识元，而是基于研究人员相互引用而形成的知识之间、人员之间的相互联络。做法相对简单，但数据的专指性不高
《2017 中国教育研究前沿与热点年度报告》；联合国环境署发布的前沿报告	取得了新的进展，获得了许多新成果，出现了新问题等	热度分析法	基于社交网络关注度或新闻报道频率、用户浏览访问或下载使用量等数据，进行综述或解析而形成报告	类似文献综述或热门盘点，比较常见的是在一些公共领域例如环境、新技术领域中，通过社交媒体关注度、新闻报道、热门评论等发现和分析热点，有些新出现、突发关注的主题也会具有一些前瞻性或预见性，有前沿的特征
各领域的相关研究论文如《国内农业经济研究 40 年：热点主题与研究前沿——基于高被引论文关键词大数据的分析》	高频关键词能较好地反映一定时期某一领域内研究者注意力的汇聚，体现这一领域的研究热点	关键词分析法	基于定量考察和可视化分析发现研究领域和发展趋势，包括词频分析、共词分析、突变/突发/爆发词检测法、基于非相关文献的知识发现法、基于概率的主题模型方法[17,39-40]等	可视化效果好，容易受到各个领域的欢迎。但要借助专门的统计分析或可视化软件实现，受软件功能限制；常常需要配合专家解读才能使报告具有可读性、易于用户理解
人工智能(AI)前沿、欧盟网络和信息安全局发布的《隐私和数据保护的设计——从政策到工程》前沿报告	热点，如 AI 是研究热点	成果追踪法	通过对最新的研究成果、基金项目、课题、学术会议、专利或技术等进行追踪分析，而形成的具有一定前瞻性价值的研究前沿报告	成果追踪法的应用领域相对狭窄，例如利用专利追踪的只适合一些较小、较专的技术领域；利用课题或基金项目等追踪的只适合一些特定时期的热门研究，难以求全，并不适合单独用来形成学科研究前沿报告

通过对前沿分析方法和实践的调研发现：前沿的认识不统一，学科前沿的概念亟待厘清；已有的前沿分析方法由于采用数据源不全面，存在一定的不足；由专业机构或咨询公司发布的前沿报告未能针对特定目标用户，并非需求驱动。图书馆如何针对用户需求，采用更有效的方法融合多源数据，以更全面、更有前瞻性地识别学科前沿，值得深入探讨。

2 学科前沿分析的方法与框架

2.1 学科前沿的基本概念

从理论上看，学科前沿这一概念的内涵是不断变化的。而且在学术界，学科前沿经常与研究前沿、科学

ChinaXiv合作期刊

前沿、新兴主题、新兴趋势、研究热点等词语交替使用。而从实践上来看,即对各个机构的前沿分析报告进行解读,发现在常见的学科前沿研究方法或相关的分析报告中,可以发现关于"前沿"的不同的定义,这其中比较容易发现热点文献、热度文献、高被引文献的影响,可见人们在使用"前沿"这个词汇时并未深入区分前沿的内涵,常常是热点与前沿混杂出现。这些概念是否有区别?

仅以ESI数据库中热点论文和研究前沿的定义为例——热点论文(Hot Paper)是指在最近2年内发表、并且在最近2个月内被引用次数达到某一学科领域的前0.1%之列的论文;研究前沿(Research Fronts)是通过共被引关系聚类出的一组高被引论文,再通过对聚类中论文题目的分析而形成。虽然是科睿唯安独有的体系和评判标准,但仍然可以看出热点和前沿有很明显的区别,首先,热点论文具有更清晰的时间敏感性;其次,热点论文有明确的阈值且阈值的设定与学科领域密切相关。相比较而言,研究前沿的时间跨度是10年,每个前沿的核心论文数2-50篇不等,都是高被引或热点论文,因为是基于共被引关系聚类出来的,虽也有明确的学科领域区分,但其实学科交叉性比较明显,更多体现的是科学研究之间存在的某种特定的共性[32]。

再结合对一些前沿分析报告(包括上文提到的几种前沿报告)的研读可以发现,学科热点和学科前沿确实存在明显的区别,前者重点基于"热度",是对现有或已有的"高显示度"成果的揭示;后者重点在于"关联",通过追踪分析研究的脉络发现具有潜力或前瞻性的成果。

笔者认为,研究前沿、科学前沿、学科前沿、新兴主题、新兴趋势、研究热点、学科热点、热点论文等这些词语的关系如图1所示。

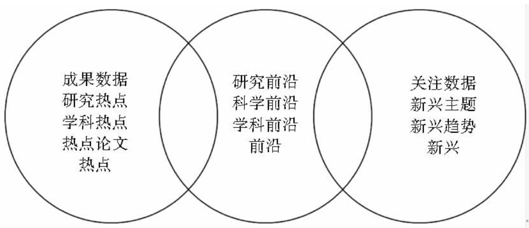


图1 研究前沿相关概念辨析

综上所述,研究前沿、学科前沿与科学前沿概念内涵相同,新兴主题与新兴趋势概念内涵相同,研究热点、热点论文与学科热点的内涵相同。研究前沿是科学研究中最新、最先进和最有发展潜力的研究主题,学科前沿是表征各学科或跨学科领域最具有影响力的、具有战略规划价值、代表未来研究方向的研究主题和关键思想。其本质是在某一个研究领域内处于领先位置的成果和思想。学科前沿可以是研究热点中那些虽然已经有很多的研究,但仍然具有发展潜力和影响力的主题,也可能是新兴趋势里那些具有战略规划价值,成为未来研究方向的主题或思想。

2.2 学科前沿分析方法的基本要素

由于对学科前沿这一概念的内涵理解不同,导致了不同学者和机构在析出学科前沿时采用不同的分析方法。在前文所述的概念基础上,学科前沿的研制方法应具备以下基本要素即基本条件,学科前沿的分析框架应从不同的角度满足这些基本条件。

(1)基于多源数据进行分析。大数据环境下前沿分析需要涵盖多种学科数据,包括期刊、会议、专利、图书、科研课题或项目/基金等。既包括学科已有的成果数据,也要包括用户使用的数据。

(2)多种分析方法综合运用。前沿分析方法的主要利弊对比见上文1.2小节的分析。可以看出,任何一种方法都有其优势和局限性。要全面析出和描述学科前沿,需要综合运用各种分析方法,最大限度发挥各种方法优势,弥补其不足。

(3)定量与定性分析相结合。学科前沿分析报告不仅要依赖文献计量分析的结果,还要充分结合专家智慧,流传不息的、权威的前沿分析报告中的每个前沿都经由强大的专家团队进行专业解析,一份只有数据却没有灵魂的报告是不会长远的,学科专家的智慧是前沿分析不可或缺的助力。

(4)开展前瞻性分析、指导学科布局。学科前沿分析应该具有预见性,应对学科研究的未来发展态势做出动态的、具有一定前瞻性的分析。

2.3 学科前沿分析方法的框架

根据学科前沿的概念和基本要素,可以得出学科前沿分析方法的总体框架(见图2):①数据源,包括期刊论文、专利成果、用户关注的数据、基金数据等;②分析方法,各种方法分析的结果相互印证和补充,以期对研究前沿提供一个全景式的描述;③析出的结果,包括从科研成果数据析出的研究热点,也包括专家根据自己的知识和文献综述析出的研究前沿,还包括虽然研究成果数量不多但是颇受用户关注的一些学科趋势,前两者是对已知研究重点领域进行回顾,后者是挖掘遴选潜在前沿领域;④由学者专家对析出结果进行分

chinaXiv:202304.00133v1

ChinaXiv合作期刊

析，识别出研究热点中那些虽然已经有很多的研究，但仍然具有发展潜力和影响力的主题，与新兴趋势里析出的具有战略规划价值、有可能成为未来研究方向的主题或思想的领域形成学科前沿，并对学科前沿进行解读。

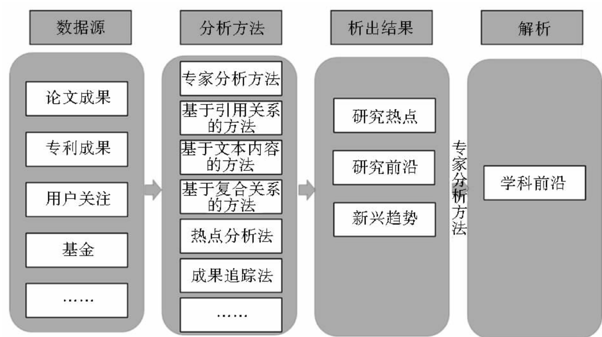


图 2 学科前沿分析方法框架

3 学科前沿分析的实施与成效

在“双一流”建设的背景和需求驱动下，中国高校图书馆率先开始学科前沿分析的方法实践，并取得了较好的成效，近年来尤以北京大学图书馆发布的两个前沿研究报告最具有代表性。北京大学图书馆于2019 年完成并发布的《北京大学科学研究前沿（2018 年版）》和《中国高校研究前沿》梳理了中国高校和北京大学各学科前沿的现状和分布，明确中国高校和北京大学各学科具有竞争优势的研究领域和研究方向，评估和判断中国高校和北京大学各学科研究的未来发展态势，拟找到实现引跑和领跑的可能发力点，为中国高校和北京大学“双一流”学科建设的布局及资源投入提供参考。下面重点以这两份报告为案例介绍学科前沿分析的特色、创新价值与成效。

3.1 学科前沿分析的特色

两个前沿分析报告均具有前沿分析研究的鲜明特点——首先是前瞻性，两个前沿报告均以“双一流”建设为目标，从学科研究态势的角度拟为中国高校和北京大学各学科的建设发展确立风向标；其次是多维度，两个前沿报告分别从文献数量、文献内容（关键词）、文献影响力（引用或下载）、用户关注、用户搜索、基金数据分析、显著性指数、平均出版年、发文或指数增长、机构分布等不同层面、不同指标的维度进行了每个研究领域的前沿分析，充分体现了前沿分析的多元特征；最后是定量与定性分析的结合，两个前沿报告不仅基于文献计量分析的多个维度进行分析，还邀请学科专家对各学科前沿的核心文献进行专业的解析，学科专家的参与是前沿研究的质量保障之一。

两个前沿分析报告综合采用了前沿分析研究的各种方法，《北京大学科学研究前沿》分别采用了引文分析法、热度分析法、关键词分析法和成果追踪法，以北京大学学者发表在中外文期刊中的高被引文献和以北京大学学者下载量最多的前 1 000 篇文献为依据进行关键词词频和共现关系分析，以北京大学用户在图书馆搜索系统中的热门搜索词为依据进行热度分析，以国家自然科学基金项目数据为依据进行成果追踪分析等；《中国高校研究前沿》分别采用了引文分析法、关键词分析法和热度分析法，以中国高校在 ESI 和 SciVal-TOP 中的高被引聚类文献为依据发现各学科前沿，以各学科各个前沿的核心文献为依据进行关键词词频分析，以各高校各学科前沿的参与度为依据进行机构的热度分析。

3.2 学科前沿分析的创新价值

两个前沿分析报告分别针对特定的目标客户，有明确的研究需求。两个报告均选取了 12 个学科领域，对这些学科领域在特定范围内的研究表现进行多数据源、多分析维度、多种分析方法的前瞻性研究，形成系统的前沿分析报告。它们的创新价值主要体现在如下方面：

（1）面向高校学科发展、精准实施内容服务。已有的前沿分析报告多是面对特定的研究领域，或是覆盖面很广的大研究领域，而非明确的学科，北京大学图书馆的学科前沿分析报告则以高校的学科发展为精准对标，致力于大学学科研究方向的前瞻性分析与预测。与近年来高校图书馆普遍开展的基于文献计量分析的学科情报研究、学科评估、学科研究支持等服务相比，学科前沿分析不仅展示了核心论文数量、被引频次和平均出版年等文献计量分析结果，而且邀请专家对每个学科领域遴选出的前沿尝试深入分析和解读。内容解析通过对每个研究前沿的全部或部分高被引核心论文的题名、关键词和摘要进行综述而形成，并由学科专家进行校验，这种定量与定性分析相结合的方式，深入到学科科研的内容本身，对于学科发展具有更明确的价值和参考意义。

（2）整合数据源、指标体系和分析维度，建立综合分析模型。文献调研和实践显示已有前沿分析报告多基于单一数据源或单一分析方法，北京大学图书馆的两个分析报告的数据源，从来源种类来看分别包含商业数据库中的研究成果、目标机构的用户行为数据（包括下载和搜索等行为）；从来源渠道来看分别包含美国

chinaXiv:202304.00133v1

科睿唯安公司 Web of Science 核心合集数据库和 ESI 数据库、荷兰爱思唯尔公司 Scopus 数据库、SciVal 数据库和 Elsevier 电子期刊、中国知网学术期刊网络出版总库等；从来源数据类型来看既包括期刊论文、会议论文等文献信息，也包含关键词、热搜词、热点描述等文本信息，还包括引文篇次、机构数等数字信息。

在指标体系方面，选取多种来源数据中的多个指标，并且对其中部分指标进行重新定义，例如重新定义了 ESI 研究前沿的数据范畴，将 ESI 的研究前沿（Researh Fronts，RF）扩展为新兴研究领域（Emerging Research Area，ERA）。ERA 具体的标准有 3 个：①必须与足够数量的 RF 论文共同引用——根据 RF 的大小，这个限制从 2 篇到 12 篇不等；②必须有足够数量的论文引用，同时也引用 RF 论文，这一限制目前被设置为共同引用阈值的两倍；③与任何一篇 RF 论文的最大余弦相似度必须超过 0.05。在扩展的文献范围（ERA）内分别依据中国高校参与度、中国高校发文量、全球总被引量、全球篇均被引等标准筛选出中国高校各学科的前 20－25 个研究前沿。对于 SciVal，报告针对学科文献产出量筛选出显著性指数（Prominence_Percentile，PP）值大于 99% 或 90%、发文量全球前 50% 的 Topic，并分别依据发文量、总被引量、显著性指数、发文量增长、显著性指数增长等标准筛选出中国高校各学科的前 20－25 个研究前沿。分析报告不仅对目标机构的科研论文、基金课题、学者关注等多角度进行研究热点的解析，而且使用目标机构用户搜索热词（见图 3）进行验证和对比，突破以往科学前沿分析仅仅从论文发表及被引情况进行单一指标判定的做法，评判标准科学、客观、全面，具有较高的前瞻性和参考度。

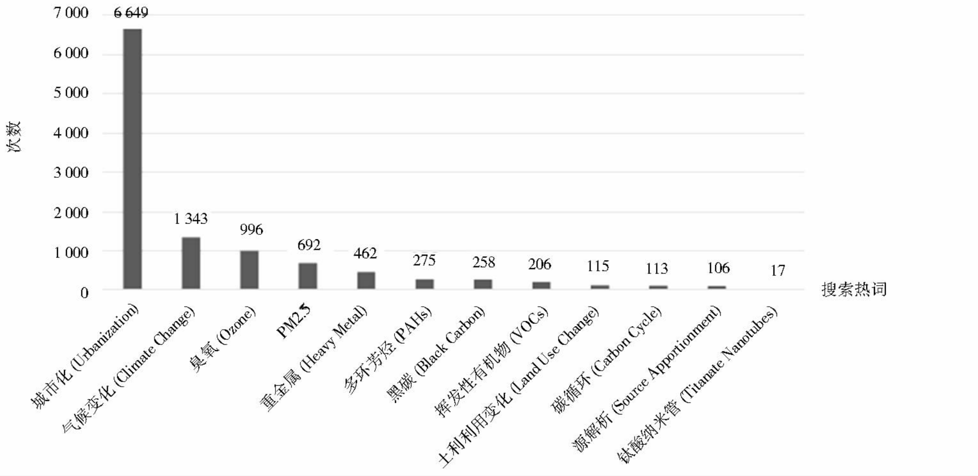


图 3 《北京大学科学研究前沿（2018）》－环境科学与工程研究前沿：热门搜索分析

从分析维度来看，报告在各学科具体的研究前沿分析中采用了文献计量分析、关键词词频分析和内容解析等方法。文献计量计算了各学科各个研究前沿的核心论文数、总被引次数、篇均被引和平均出版年；关键词词频分析对被引频次进行加权，并生成词云图；内容解析则通过文献综述和专家校验等方式实现。

（3）宏观分析与微观分析结合的视角。报告采用了宏观分析与微观分析相结合的方法，宏观分析关注中国高校各学科的前沿的整体表现水平、在中国和全球的参与度；微观分析关注具体前沿的主题和机构分布。报告既关注中国高校各学科基于共被引关系的重点研究前沿，也关注基于显著性指数、潜力值等综合表征的新兴前沿和最具潜力前沿。宏观与微观相结合的视角使学科前沿分析既可成为学科发展的风向标，又能在明确的学科范围内提供计量性和预测性信息，对于高校图书馆更好地融入学科发展有积极的意义。

3.3 成效分析

学科前沿分析对于学科设置、学科发展以及学科研究均有指导意义。

《北京大学科学研究前沿》报告发布以后，引起北京大学校内各相关学科的高度关注，国家发展研究院（经济学、管理学）、现代农学院（经济学）、地球与空间学院（地质学、环境科学与工程）等院系单位纷纷与图书馆联络希望订制前沿报告——有的是为了更有针对性地面向二级学科订制报告、有的是为了申请建设本校没有的新学科。图书馆为院系/学科建设与规划提供了一个新的角度与思路，因而受到学院领导及相关管理决策部门的肯定与好评，因此得以与学校学科办、科研部等单位共同申请课题，合作开展研究和相关会议培训等活动，不断融入大学学术事务。前沿报告同

chinaXiv:202304.00133v1

ChinaXiv合作期刊

时也获得图书馆界同行的欢迎和认可,纷纷索取报告和表达合作交流的意愿。前沿报告的研究方法和成果等也以北京大学图书馆的名义申请或申报了相关的课题和奖项,并且已获得了一些奖励,例如高校图书馆数字资源采购联盟(DRAA)“数字资源深度利用”课题二等奖等。《中国高校研究前沿》依据中国高校学科设置与发展,结合全球两大知名科学研究数据出版商科睿唯安及爱思唯尔公司前沿研究相关数据源的学科范畴,梳理和分析了中国高校 12 个学科领域的研究前沿,从前沿角度把握中国高校的学科态势,为明确学科竞争优势、评判和预测学科发展方向提供了新的视角。两大知名数据源覆盖中国高校代表性文献,析出的各学科研究前沿非常广泛和均衡,既有重叠又有互补——基于 ESI 的研究前沿主要是各学科最近几年密集出现的研究领域,范畴相对集中;基于 SciVal 的研究前沿既有“新贵”、也有“世家”,既有非常新的交叉性极强的研究领域,也有非常传统的、文献特别富集的研究领域。《中国高校研究前沿》拟从 2018 年起逐年发布,将成为中国高校学科发展态势的风向标。

4 问题与展望

4.1 数据及其来源受限、数据处理困难

为实现前瞻性研究,北京大学图书馆尝试使用了用户关注数据和用户搜索数据,但这两个数据一个受限于数据提供商,在提供时间、数据格式和数据灵活性等方面都有较多困难;另一个受限于搜索平台,平台本身的更新以及日志数据的不完全、不完善等导致数据的可用性较差。再者由于数据源本身的限制以及学科的特点等,有些报告无法使用中文数据,多数报告没有涉及人文和社会科学学科。

不同数据源的归并去重难以完成,这是由于不同数据的年代范围、文献体量、聚类算法等都有较大差异,同时由于分析的基础数据是由出版商所提供,数据内容格式等也都受到限制。

4.2 分析技术和工具受限

为了挖掘热点与前沿,北京大学图书馆重点使用了 SPSS 和 CiteSpace 等工具,但这些工具的分析能力和效果均与数据以及分析人员的技术水平相关,由于数据源的限制只进行了关键词级别的分析,对于关键词共现关系的分析也只开展到“点度中心性”的层面。由于没有专业的技术团队和技术手段做海量数据的归并、清洗和关联、挖掘等,得出的分析结论常常是基于离散的关键词的堆积,据此形成的前沿常常是比较宽泛的领域,不能充分体现前沿的特质。

4.3 缺乏专业的前沿内涵解析

前沿解析是一个复杂而庞大的工程,需要组成各学科领域的专家团队才能完成,需要建立有效的长期的组织机制。

为了解决这些问题,需要在前沿分析工作中尝试引入更多的数据源,与相关机构的技术团队及学科专家深度合作,共同挖掘和发现学科前沿。在数据处理方面需要采用自建数据库将所有基础分析数据进行获取、保存、清洗、规范化处理等,再定义自己的算法、运用数据挖掘和数据分析技术进行自主分析。当然也需要建立稳定的、长效的学科专家团队。

5 结语

图书馆开展学科研究前沿分析服务是支撑学科建设的重要举措之一,图书馆拥有海量信息资源、相关的分析工具、文献计量学及其他分析方法、学科服务和相关服务基础等,在开展前瞻性的学科前沿研究方面具有天然的优势。前沿研究目前更适合理工学科,尤其是新兴和交叉的学科领域,但随着大数据时代的到来,数据化思维对于人文和社会科学领域的影响也越来越大,人文社会科学领域也越来越关注学科前沿。以北京大学现代农学院(筹)为例,目前该学院拟新设置农林经济管理学科,在设置之前就需要对该学科的研究热点和前沿进行分析对比,并了解学校科研人员在该学科领域的学术表现等。学科前沿分析在大学有清晰的应用前景。

前沿研究分析目前存在着数据和技术瓶颈,从数据的角度来看,完整全面地获取数据并对不同来源的数据进行去重、归并、清洗、规范等,全面追踪用户使用信息的行为数据并对这些数据进行准确的归类和判断等,都还需要更多的数据源以及技术支撑;从技术的角度来看,前沿研究的前瞻性和预测性不足,现有技术对于后评估方法、已发生的用户使用行为有较为成熟的分析处理方案,但对于前评估方法、即将发生的用户使用行为尚未形成较好的处置手段。未来图书馆的学科研究前沿分析应与数据公司、技术公司深度合作,并可反过来推进数据的建设,尤其是人文社会科学各种科研分析评价数据,如图书专著、艺术作品等信息源及其影响力等数据的建设。此外,还应该推进数据公开,对前沿分析报告的后台数据提供开放获取的渠道,方便查询和验证,让更多人能够发现和利用前沿文献。

chinaXiv:202304.00233v1

chinaXiv:202304.00133v1

参考文献：

[1] 习近平. 决胜全面建成小康社会,夺取新时代中国特色社会主义伟大胜利——在中国共产党第十九次全国代表大会上的报告[EB/OL]. [2020 - 05 - 12]. http://www.12371.cn/2017/10/27/ARTI1509103656574313.shtml.

[2] 国务院. 国务院关于印发统筹推进世界一流大学和一流学科建设总体方案的通知[EB/OL]. [2020 - 05 - 12]. http://www.gov.cn/zhengce/content/2015-11/05/content_10269.htm.

[3] PRICE D J. Networks of scientific papers[J]. Science,1965, 149(3683): 510 - 515.

[4] SMALL H. Co-citation in the scientific literature: a new measure of the relationship between two documents[J]. Journal of the American Society for Information Science,1973,24(4): 265 - 269.

[5] PERSSON O. The intellectual base and research fronts of JASIS 1986 - 1990[J]. Journal of the American Society for Information Science, 1994,45(1): 31 - 38.

[6] GARFIELD E. Research fronts[J]. Current contents, 1994(41): 3 - 7.

[7] MORRIS S A, YEN G, WU Z, et al. Time line visualization of research fronts[J]. Journal of the American Society for Information Science and Technology,2003, 54(5): 413 - 422.

[8] KLEINBERG J. Bursty and hierarchical structure in streams[J]. Data mining and knowledge discovery, 2003,7(4):373 - 397.

[9] CHEN C. Citespace II: Detecting and visualizing emerging trends and transient patterns in scientific literature[J]. Journal of the American Society for Information Science and Technology,2006,57(3): 359 - 377.

[10] 黄晓斌,吴高. 学科领域研究前沿探测方法研究述评[J]. 情报学报,2019(8):872 - 880.

[11] 陈仕吉. 科学研究前沿探测方法综述[J]. 现代图书情报技术,2009(9):28 - 33.

[12] 张惠荣, 姜玉梅,何冰. 基于前沿和热点的学科门户信息服务模式研究——以天津医科大学肿瘤医院图书馆为例[J]. 图书馆工作与研究,2010(6):85 - 87.

[13] LAW J, BAUIN S, COURTIAL J P, et al. Policy and the mapping of scientific change: a co-word analysis of research into environmental acidification[J]. Scientometrics,1988, 14(3/4): 251 - 264.

[14] ARIS A, SHNEIDERMANE B, QAZVINIAN V, et al. Visual overviews for discovering key papers and influences across research fronts[J]. Journal of the American Society for Information Science and Technology,2009, 60(11): 2219 - 2228.

[15] BRAAM R, WEB H, VAN RAAN. Mapping of science by combined co-citation and word analysis Ⅱ: dynamical aspects[J]. Journal of the American Society for Information Science,1991(4): 252 - 264.

[16] VAN DEN BESSELAAR P, HEIMERIKS G. Mapping research topics using word-reference co-occurrences: a method and an exploratory case study[J]. Scientometrics,2006,(3): 377 - 393.

[17] 郑彦宁,许晓阳,刘志辉. 基于关键词共现的研究前沿识别方法研究[J]. 图书情报工作, 2016,60(4):85 - 92.

[18] LI M, CHU Y. Explore the research front of a specific research theme based on a novel technique of enhanced co-word analysis [J]. Journal of information science, 2017, 43(6):725 - 741.

[19] ZHANG T, CUI H, OUYANG Z. Detecting research focus and research fronts in the medical big data field using co-word and co-citation analysis[C] // 2018 IEEE 20th international conference on high performance computing and communications; IEEE 16th international conference on smart city; IEEE 4th international conference on data science and Systems (HPCC/SmartCity/DSS). New York:IEEE,2018:313 - 320.

[20] HOU J, YANG X, CHEN C. Emerging trends and new developments in information science: a document co-citation analysis (2009 - 2016)[J]. Scientometrics,2018,115(2):869 - 892.

[21] 梁丽,谢凤杰,池丽旭,等. 特定学科热点和前沿主题研究方法实证分析[J]. 图书馆杂志, 2018, 37(1):9 - 26, 32.

[22] ARYADOUST V, ANG B. Exploring the frontiers of eye tracking research in language studies: a novel co-citation scientometric review[J]. Computer assisted language learning, 2019,32(8):1 - 36.

[23] 陈雪,黄奇. 基于共词分析的国内图书情报领域本体研究热点探析[J]. 图书馆学研究, 2019(8):2 - 8.

[24] PARK I, LEE K, YOON B. Exploring promising research frontiers based on knowledge maps in the solar cell technology field[J]. Sustainability, 2015, 7(10): 13660 - 13689.

[25] PARK I, YOON B. Identifying promising research frontiers of pattern recognition through bibliometric analysis[J]. Sustainability, 2018, 10(11). DOI: 10.3390/su10114055.

[26] 许晓阳,郑彦宁,刘志辉. 论文和专利相结合的研究前沿识别方法研究[J]. 图书情报工作,2016,60(24):97 - 106.

[27] 白如江,冷伏海,廖君华. 一种基于多数据源主题对比的科学研究前沿识别方法[J]. 情报理论与实践, 2017, 40(8): 43 - 48, 36.

[28] 刘博文,白如江,周彦廷,等. 基金项目数据和论文数据融合视角下科学研究前沿主题识别——以碳纳米管领域为例[J]. 数据分析与知识发现,2019(8):114 - 122.

[29] 孙震. 基于科学论文多源数据的研究前沿集成识别模型研究[J]. 情报杂志, 2016,35(8):95 - 100.

[30] 张维冲,王芳,赵洪. 多源信息融合用于新兴技术发展趋势识别 - 以区块链为例[J]. 情报学报, 2019,38(11):1166 - 1176.

[31] 谭晓,李辉. 基于多源数据知识融合方法的研究前沿识别[J]. 现代情报,2019,39(8):29 - 36.

[32] 中国科学院科技战略咨询研究院,中国科学院文献情报中心,科睿唯安中国办公室. 2018 研究前沿[EB/OL]. [2020 - 04 - 02]. https://clarivate.com.cn/blog/2018researchfronts/.

[33] 蔡方,吉久明,蒋志强,等. 2018 年全球工程前沿简介[J]. 工程,2018(6):748 - 753.

ChinaXiv合作期刊

[34] Notes from the AI frontier-modeling the impact of AI on the world economy[EB/OL]. [2020 - 04 - 02]. https://www.mckinsey.com/featured-insights/artificial-intelligence/notes-from-the-frontier-modeling-the-impact-of-ai-on-the-world-economy#0.
[35] Frontiers 2017: emerging issues of environmental concern[EB/OL]. [2020 - 04 - 02]. https://www.unenvironment.org/resources/frontiers-2017-emerging-issues-environmental-concern.
[36] Frontiers 2018/19——emerging issues of environmental concern. [EB/OL]. [2020 - 04 - 02]. https://wedocs.unep.org/bitstream/handle/20.500.11822/27538/Frontiers1819.pdf? sequence = 1&isAllowed = y.
[37] 科睿唯安公司. ESI 快速使用指南[EB/OL]. [2020 - 04 - 02]. http://dbnav.lib.pku.edu.cn/sites/default/files/ESI.pdf.
[38] 北京大学图书馆. 北京大学科学研究前沿(2018 版)[EB/OL]. [2020 - 04 - 02]. https://www.lib.pku.edu.cn/portal/fw/kyzc/zhishichanquan.
[39] 李锋. 探寻学科前沿的四种研究路径——以数字阅读为例[J]. 图书馆理论与实践,2017(12):5 - 9.
[40] 白如江,冷伏海,廖君华. 科学研究前沿探测主要方法比较与发展趋势研究[J]. 情报理论与实践,2017(5):33 - 38.
[41] 赵蓉英,许丽敏. 文献计量学发展演进与研究前沿的知识图谱探析[J]. 中国图书馆学报,2010(5):60 - 68.
[42] 王效岳,刘自强,白如江,等. 基于基金项目数据的研究前沿主题探测方法[J]. 图书情报工作,2017,61(13):87 - 98.
[43] 张晖,杨小彦,赵旭剑,等. 基于社会网络关注度的学科前沿热点挖掘[J]. 郑州大学学报(理学版),2018(3):46 - 52.
[44] 刘 毅,林世爵. 基于高价值专利分析的移动互联网产业前沿技术跟踪研究[J]. 科技创新发展战略研究,2016(6):32 - 37.
[45] 教育研究杂志社. 2017 中国教育研究前沿与热点年度报告[N]. 中国教育报,2018 - 04 - 12(6).

作者贡献说明:
张春红:文章主体框架设计、案例分析和结语;
肖珑:文章框架修订、文字修订;
贺飞:文章主旨定位、研究的价值挖掘;
唐勇:文章框架调整、文献综述、研究方法对比分析。

chinaXiv:202304.00133v1

The Study of Methods and Practices of Identifying Discipline Research Fronts ——Taking Peking University Library as an Example

Zhang Chunhong[1]　Xiao Long[1]　He Fei[2]　Tang Yong[1]
[1] Peking University Library, Beijing 100871
[2] Office of Academic Development, Peking University, Beijing 100871

Abstract: [**Purpose/significance**] By reviewing and analyzing the methods and practices, this paper summarized the effects, advantages and disadvantages of the identifying discipline research fronts. This paper clarified the concept of the discipline research fronts and explored the development of frontier analysis services based on multi-source data in libraries. [**Method/process**] Using literature study and cases study ,this paper explored the studies of identifying discipline research fronts by comparative analysis. Then it took two cases of identifying discipline research fronts of Peking University library as examples, explores the connotation, characteristics, indexes and research methods of discipline research frontier analysis service in university library, which provides reference for the development of the services. [**Result/conclusion**] The results show that the accurate benchmarking of discipline development needs, diversity of data sources, comprehensive analysis of the source data, and comprehensive application of various methods have been significantly improved the effects of the frontier analysis service in the aspects of clarifying the research fields and research directions of each discipline. Which can help to clearly understand, evaluate and judge the future development trend of the disciplines, and provide a useful reference for the layout and resource investment of the disciplines development.

Keywords: frontier analysis　frontier report　discipline research frontier　hotspot analysis　co-word analysis